BEOORDELINGSKADER ZORGVRAAGZWAARTE

In dit document geven we een beoordelingskader voor de beoordeling van de zorgvraagzwaarte-indicator. Dit beoordelingskader is gebaseerd op de resultaten van de besprekingen in de focusgroepen en in de werkgroep zorgvraagzwaarte.

Het beoordelingskader is opgebouwd uit 5 beoordelingscriteria op basis waarvan een eindoordeel over het gebruik van de zorgvraagzwaarte-indicator per aanbieder voor de materiële controles en de contractering kan worden gemaakt.

Voor elk beoordelingscriterium is een stoplicht ontwikkeld aan de hand waarvan een oordeel kan worden geveld of het gebruik van de zorgvraagzwaarte-indicator op basis van dit beoordelingscriterium geschikt is. Steeds is naast het stoplicht aangegeven wat de criteria zijn voor de overgang van rood naar oranje, en van oranje naar groen.

In hoofdstuk 1 geven we het beoordelingskader voor de individuele beoordelingscriteria. Hoofdstuk 2 bevat de beschrijving hoe op grond van deze individuele beoordelingscriteria tot een eindoordeel kan worden gekomen. In bijlage 1 beschrijven we de statistische methoden.

1 Beoordeling op vijf beoordelingscriteria

1.1 Inhoudsvaliditeit

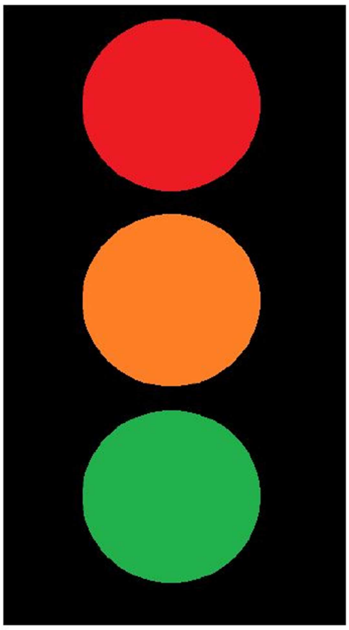
Met de inhoudsvaliditeit toetsen we of de indicator aansluit bij de beoogde doelstelling. Dit onderzoeken we door na te gaan in hoeverre de zorgvraagzwaarte-indicator de verwachte behandelinzet kan voorspellen. Als de zorgvraagzwaarte-indicator de behandelinzet onvoldoende verklaart dan is de zorgvraagzwaarte-indicator in de huidige setting niet geschikt om te gebruiken bij de contractering en de materiële controles .

Inhoudsvaliditeit meten we aan de hand van de R-kwadraat. De R-kwadraat geeft weer welk deel van de variatie in de behandelinzet kan worden verklaard door middel van verschillen in zorgvraagzwaarte en hoofdgroepdiagnose[1]. Omdat we het model van zorgvraagzwaarte niet op individueel niveau maar op groepsniveau (zorgaanbiederniveau)

[1] Mogelijk kunnen hier in een later stadium andere kenmerken aan toegevoegd worden.

wordt toegepast, berekenen we ook de R-kwadraat op groepsniveau. (Zie bijlage 1 voor de berekening van de R-kwadraat op groepsniveau).

De R-kwadraat stellen we vast voor het gehele model. De waarde van dit stoplicht geeft dus weer in hoeverre het model van zorgvraagzwaarte toepasbaar is voor alle aanbieders.



Als de R-kwadraat op groepsniveau van een model waarin de behandelinzet verklaard wordt door zorgvraagzwaarte en hoofdgroepdiagnose significant beter is dan de R-kwadraat waarin alleen hoofdgroepdiagnose als verklarende variabele wordt gebruik is de score **oranje**[1]**.**

Als de R-kwadraat op groepsniveau hoger is dan **50%**[2] [dit percentage moet beleidsmatig worden vastgesteld] van de maximaal haalbare R-kwadraat[3] dan is de score **groen.**

[1] Deze grenswaarde kan worden vastgesteld door een empirische analyse met behulp van een F-toets. Aan de hand van de werkelijke gegevens kan een simulatie worden uitgevoerd, waarin de minimale R-kwadraat wordt vastgesteld.

[2] Bij een R-kwadraat hoger dan 50% wordt het merendeel van de variatie verklaard door het model. Dit is een gebruikelijke grenswaarde in de praktijk. Bij een R-kwadraat hoger dan 50% van de maximaal haalbare R-kwadraat wordt het merendeel van de mogelijk te verklaren variantie verklaard.

[3] Deze grenswaarde kan worden vastgesteld door een analyse waarbij een regressie wordt uitgevoerd, waarin ook de regio en de behandelinzet in het verleden als variabelen worden toegevoegd aan het regressiemodel. Voor het vaststellen van deze grenswaarde is het noodzakelijk om een analyses uit te voeren op basis van gegevens van het DIS.

1.2 Registratievergelijkbaarheid

Met registratievergelijkbaarheid toetsen we of de gehanteerde gegevens goed geregistreerd zijn. Hiervoor kunnen we slechts een kwalitatieve toets uitvoeren. In deze

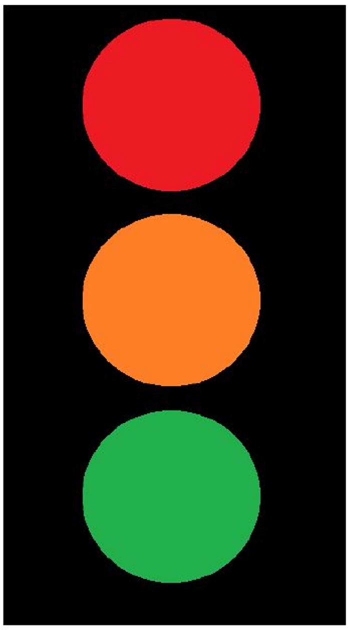
kwalitatieve toets gaan we na of het proces van registreren aan bepaalde voorwaarden voldoet. In beginsel zijn er drie voorwaarden:

1. De gegevens worden door een andere entiteit gecontroleerd dan de entiteit die registreert.

2. De gegevens worden op individueel niveau vastgelegd.

3. De gegevens worden gebruikt in het primaire proces van zorgverlening.

De toetsing van dit criterium vindt plaats op het niveau van de onderliggende gegevens voor het vaststellen van de zorgvraagzwaarte-indicator. Voor de zorgvraagzwaarte-indicator 1.0 betreft dit daarmee de aard van de stoornis, de nevendiagnose, psychosociale factoren en de GAF score.



Als voor alle gebruikte gegevens (van de onderliggende factoren) geldt dat aan minimaal **1** van de bovenstaande voorwaarden wordt voldaan is de score **oranje.**

Als voor alle gebruikte gegevens (van de onderliggende factoren) geldt dat aan minimaal **2** van de bovenstaande voorwaarden wordt voldaan is de score **groen.**

1.3 Populatievergelijkbaarheid

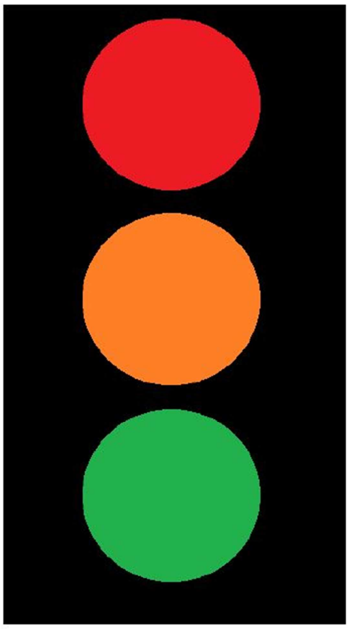
Bij de populatievergelijkbaarheid toetsen we of de zorgvraagzwaarte-indicator in afdoende mate gecorrigeerd wordt voor verschillen in de populatie. Als dit niet het geval is dan kan geen adequate vergelijking tussen aanbieders plaatsvinden. Aangezien dat wel de manier is om in de contractering afspraken te maken, kan bij een onvoldoende populatievergelijkbaarheid geen gebruik worden gemaakt van de zorgvraagzwaarte-indicator. De beoordeling van de populatievergelijkbaarheid vindt plaats op het niveau van het model.

Voor het toetsen van de populatievergelijkbaarheid toetsen we of de zorgvraagzwaarte-indicator adequaat gecorrigeerd wordt voor de kenmerken die bij het vaststellen van de zorgvraagzwaarte-indicator 1.0 zijn vastgesteld. Dit zijn:

- de aard van de stoornis
- de ernst van de stoornis (Symptomatologie)
- de ernst van de stoornis (kwaliteit van leven)
- psychosociale factoren
- nevendiagnoses
- het beloop van de stoornis

Het vergt een inhoudelijke afweging of de correctie die bij het vaststellen van de zorgvraagzwaarte-indicator wordt toegepast adequaat is. Voor deze afweging stellen deskundigen per kenmerk vast of de correctie adequaat is.

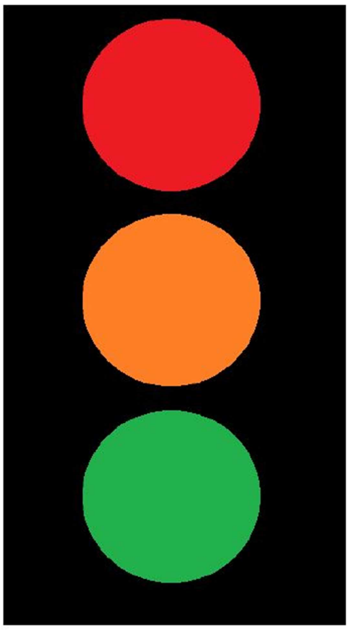
Daarnaast moeten de deskundigen zich buigen over de vraag of bovenstaande lijst nog juist is, of dat nieuwe inzichten er toe leiden dat aanpassingen aan de lijst met kenmerken waarvoor gecorrigeerd moet worden gewenst is.



Bij adequate correctie voor minimaal **2/3** [Dit aandeel moet door deskundigen worden vastgesteld] van de kenmerken, zoals hierboven vermeld is de score **oranje**
Bij adequate correctie voor **alle** kenmerken is de score **groen.**

1.4 Statistische betrouwbaarheid

De statistische betrouwbaarheid wordt gemeten op het niveau van de aanbieder. Voor elke aanbieder stellen we vast of de aanbieder voldoende patiënten behandeld per klasse van zorgvraagzwaarte en hoofdgroepdiagnose om statistisch betrouwbaar de zorgvraagzwaarte-klasse vast te kunnen stellen.



Als in minimaal **50%** van de groepen van zorgvraagzwaarte en hoofdgroepdiagnose de aanbieder ofwel geen ofwel voldoende patiënten[3] heeft is de score **oranje**.

Als in minimaal **80%** van de groepen van zorgvraagzwaarte en hoofdgroepdiagnose de aanbieder ofwel geen ofwel voldoende patiënten[3] heeft, is de score **groen**

[3] Wat voldoende is, kan bepaald worden aan de hand van een poweranalyse. In de poweranalyse maken we gebruik van:

- de variatie in behandelinzet binnen de combinatie van zorgvraagzwaarte-klasse en hoofdgroepdiagnose *en*
- het aantal waarnemingen binnen deze groep *en*
- het verschil in verwachte behandelinzet met de naaste klassen van zorgvraagzwaarte van dezelfde hoofddiagnosegroep.

Voor het vaststellen van deze grenswaarde is het noodzakelijk om een analyse uit te voeren op basis van gegevens van het DIS.

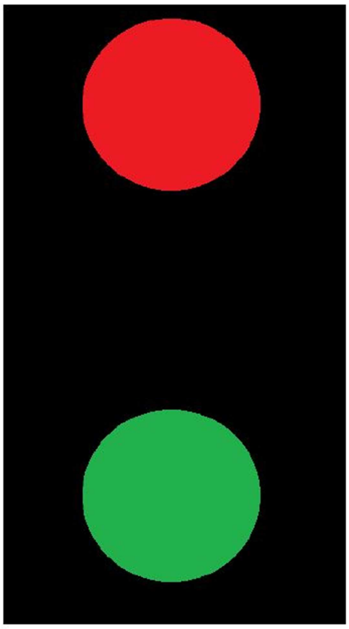
1.5 Representativiteit

Het laatste criterium is de representativiteit van de gegevens. De representativiteit is van belang op het moment dat voor het vaststellen van de zorgvraagzwaarte-indicator gebruik is gemaakt van een steekproef. Bij de representativiteit toetsen we of een kenmerk gebruikt kan worden voor het vaststellen van de zorgvraagzwaarte-indicator. Als de

gegevens niet representatief zijn kan de betreffende indicator niet gebruikt worden. Dit heeft dan vooral consequenties voor de populatievergelijkbaarheid.

De representativiteit toetsen we aan de hand van een statistische toets, waarin onderzocht wordt of de samenstelling van de gebruikte dataset overeenkomt met de totale dataset. Als uit de analyse blijkt dat de gehanteerde steekproef niet representatief is kan allereerst gezorgd worden dat de dataset alsnog representatief wordt.

Bij de beoordeling van de representativiteit wordt geen score oranje toegekend. De score op representativiteit is groen als de samenstelling van de steekproef minder dan 5% afwijkt van de totale populatie. Als dat niet het geval is, is de score rood.



Als uit een Chi-kwadraattoets[3] blijkt dat samenstelling van de steekproef minder dan **5%** afwijkt van de totale populatie dan is de score op representativiteit **groen.**

[3] De Chi-kwadraat toets geeft weer wat de kans is dat de samenstelling van de steekproef afwijkt van de samenstelling van de gehele populatie. Een kans van 5% betekent hiermee dat het 95% zeker is dat de samenstelling van de steekproef overeenkomt met de samenstelling van de populatie.

Bijlage 1: statistische methoden

R-kwadraat op groepsniveau

Voor het vaststellen van de R-kwadraat op groepsniveau voeren we een regressie uit. We gaan vooralsnog uit van OLS. In een later stadium zouden ook andere regressiemethoden die nog beter aansluiten bij de verdeling van de behandelinzet over de patiënten worden getoetst.

De verklaarde variabele in de regressie is de behandelinzet in termen van minuten en dagenverblijf. Deze zijn gewogen aan de hand van de methode die DBC-onderhoud heeft uitgevoerd.

De verklarende variabelen zijn dummies van de interactietermen van zorgvraagzwaarte en hoofdgroepdiagnose. Elke dummy is 1 als de betreffende patiënt in de zorgvraagzwaarte-klasse en hoofdgroepdiagnose is ingedeeld en 0 als de betreffende patiënt niet in die groep thuishoort.

De regressie voeren we uit op het niveau van de combinatie van individuele patiënten. De R-kwadraat op groepsniveau stellen we in drie stappen vast:

1. We stellen de gemiddelde behandelinzet en de gemiddelde voorspelde behandelinzet per individu vast
2. We stellen de som vast van de gemiddelde behandelinzet en de gemiddelde voorspelde behandelinzet per groep (hoofdgroepdiagnose, zorgvraagzwaarte en zorgaanbieder)
3. De R-kwadraat is vervolgens gelijk aan de correlatie in het kwadraat tussen de som van de werkelijke behandelinzet en de verwachte behandelinzet per groep.

F-toets op de R-kwadraat (grenswaarde tussen rood en oranje)

Vervolgens voeren we een F-toets uit om te toetsen of de R-kwadraat gelijk is aan 0. De F-statistiek is hierbij gelijk aan

$$F = \frac{R^2 * (N-K-1)}{(1-R^2) * K}$$

Daarbij is N gelijk aan het aantal waarnemingen en K gelijk aan het aantal variabelen dat in de regressie is opgenomen. Deze F-toets is een F-toets op individueel niveau. Als de behandelinzet op individueel niveau een voorspellende waarde heeft dan is dat ook het

geval op groepsniveau. De F-toets op individueel niveau is daarmee een stringentere eis dan een vergelijkbare toets op groepsniveau.

De F-toets betreft vervolgens een toets met (K, N-K-1) vrijheidsgraden. Hiermee kan een p-waarde worden vastgesteld. Als deze kleiner is dan 0.05 dan is de R-kwadraat met kans 0,95 ongelijk aan 0.

Als de R-kwadraat gelijk is aan 0 dan heeft het toepassen van de zorgvraagzwaarte-indicator geen toegevoegde waarde. De zorgvraagzwaarte-indicator kan dan niet gebruikt worden.

Regressie met variabele regio (Grenswaarde tussen oranje en groen)

Voor het toepassen van de zorgvraagzwaarte-indicator is een scherpere grenswaarde benodigd. De toets tussen rood en oranje geeft immers alleen aan of de R-kwadraat toegevoegde waarde heeft boven het niet gebruiken van de zorgvraagzwaarte-indicator.

Hiervoor toetsen we of de gevonden R-kwadraat op groepsniveau hoger is dan 50% van een maximaal haalbare R-kwadraat op groepsniveau. Voor het vaststellen van de maximaal haalbare R-kwadraat voegen we twee verklarende variabelen aan het model toe:

1. De gemiddelde behandelinzet van de aanbieder in het verleden

2. Een dummy voor de regio van de aanbieder

Beide kenmerken voegen we toe als aanvullende variabelen aan het model.

Met dit alternatieve model schatten we opnieuw het regressiemodel. Het resultaat van dit model geeft een inschatting van de maximaal haalbare R-kwadraat voor een model waarin we variatie willen verklaren. Indien de gevonden R-kwadraat op groepsniveau hoger is dan 50 % van deze maximale R-kwadraat dan kan een score groen worden toegekend.

Minimaal aantal waarnemingen (grenswaarde minimaal aantal waarnemingen per groep)

Het minimaal aantal waarnemingen stellen we per zorgvraagzwaarte-klasse per hoofdgroepdiagnose vast. We voeren een t-toets uit per ZVZ-klasse of de betreffende klasse een lagere gemiddelde behandelinzet heeft dan de naasthogere klasse binnen dezelfde hoofdgroepdiagnose. Dit betreft een eenzijdige t-toets per klasse. Daarbij gaan we uit van een t-waarde van 1,64. Deze t-waarde hoort bij een power van 10%.

Deze toets voeren we uit voor alle ZVZ-klassen, waarbij we per hoofdgroepdiagnose nagaan hoeveel waarnemingen per ZVZ-klasse nodig zijn om aan te kunnen geven dat deze groep een signifcant lagere behandelinzet heeft dan de naasthogere groep. Door tegelijkertijd

ook te testen of de verzekerden in de naasthogere groep een significant hogere behandelinzet hebben kan worden nagegaan wat het aantal waarnemingen per groep is.

In formulevorm ziet dit er als volgt uit:

$$t = \frac{BHI_{ZVZ1,HDG} - BHI_{ZVZ2,HDG}}{S * \sqrt{\frac{1}{N_{ZVZ1,hgd,za}} + \frac{1}{N_{ZVZ2,hgd,za}}}}$$

,

$BHI_{ZVZ1,HDG} =$ De gemiddelde behandelinzet in zorgvraagzwaarte-klasse 1 en een specifieke Hoofddiagnosegroep

$N_{ZVZ1,HDG,za} =$ Het aantal patiënten van een zorgaanbieder in zorgvraagzwaarte-klasse 1 en een specifieke hoofddiagnosegroep

$\sigma_{ZVZ1,HDG} =$ de standaardafwijking in de gemiddelde behandelinzet in zorgvraagzwaarte-klasse 1 en een specifieke Hoofddiagnosegroep

$S =$ de gewogen standaardafwijking van de behandelinzet van patiënten in ZVZ 1 en ZVZ 2van dezelfde hoofddiagnosegroep

Een laatste stap bij het vaststellen van het minimaal aantal waarnemingen is om na te gaan bij hoeveel van de klassen een aanbieder meer patiënten heeft dan deze ondergrens of dat de aanbieder in deze groep geen patiënten behandelt.

Hierbij hanteren we een grenswaarde van 50% voor het onderscheid tussen rood en oranje en een grenswaarde van 80% tussen oranje en groen. Om een score groen te behalen moet een aanbieder dus in minimaal 80% van de groepen waar hij patiënten behandelt meer waarnemingen hebben dan de ondergrens.